

When FLUE Meets FLANG: Benchmarks and Large Pretrained Language Model for Financial Domain

Year: 2022 | Citations: 140 | Authors: Raj Sanjay Shah, Kunal Chawla, Dheeraj Eidnani

Abstract

Pre-trained language models have shown impressive performance on a variety of tasks and domains. Previous research on financial language models usually employs a generic training scheme to train standard model architectures, without completely leveraging the richness of the financial data. We propose a novel domain specific Financial LANGuage model (FLANG) which uses financial keywords and phrases for better masking, together with span boundary objective and in-filing objective. Additionally, the evaluation benchmarks in the field have been limited. To this end, we contribute the Financial Language Understanding Evaluation (FLUE), an open-source comprehensive suite of benchmarks for the financial domain. These include new benchmarks across 5 NLP tasks in financial domain as well as common benchmarks used in the previous research. Experiments on these benchmarks suggest that our model outperforms those in prior literature on a variety of NLP tasks. Our models, code and benchmark data will be made publicly available on Github and Huggingface.